Swarm Intelligence
https://doi.org/10.1007/s11721-021-00191-9

Collective preference learning in the best-of-*n* problem

From best-of-*n* to ranking *n*

Michael Crosscombe[1] · **Jonathan Lawry**[1]

Accepted: 24 April 2021

Abstract
Decentralised autonomous systems rely on distributed learning to make decisions and to collaborate in pursuit of a shared objective. For example, in swarm robotics the best-of-*n* problem is a well-known collective decision-making problem in which agents attempt to learn the best option out of *n* possible alternatives based on local feedback from the environment. This typically involves gathering information about all *n* alternatives while then systematically discarding information about all but the best option. However, for applications such as search and rescue in which learning the ranking of options is useful or crucial, best-of-*n* decision-making can be wasteful and costly. Instead, we investigate a more general distributed learning process in which agents learn a preference ordering over all of the *n* options. More specifically, we introduce a distributed rank learning algorithm based on three-valued logic. We then use agent-based simulation experiments to demonstrate the effectiveness of this model. In this context, we show that a population of agents are able to learn a total ordering over the *n* options and furthermore the learning process is robust to evidential noise. To demonstrate the practicality of our model, we restrict the communication bandwidth between the agents and show that this model is also robust to limited communications whilst outperforming a comparable probabilistic model under the same communication conditions.

Keywords Collective learning · Preference learning · Distributed decision-making · Multi-agent systems

1 Introduction and related work

In decentralised agent-based systems, we aim to develop distributed processes whereby desirable system-level behaviour emerges from the interactions between individuals at the local level (Brambilla et al. 2013). Collective decision-making broadly describes a range

✉ Michael Crosscombe
m.crosscombe@bristol.ac.uk

Jonathan Lawry
j.lawry@bristol.ac.uk

[1] Department of Engineering Mathematics, University of Bristol, Bristol, UK

Published online: 02 June 2021

of collective behaviours including consensus formation, task allocation and fault detection (Schranz et al. 2020). Consensus formation has long been studied at the intersection of mathematics and the social sciences where models of opinion diffusion were proposed to study the dynamics of iterative belief pooling (Stone 1961; DeGroot 1974; Lehrer and Wagner 1981). Such models have proven highly influential in studying how beliefs propagate through a system of agents, particularly in the context of social network analysis (Perron et al. 2009; Baronchelli 2018). Traditionally, agents operate in a closed setting with prior knowledge about some variable of contention. Then, at each time step a weighted linear combination operator is applied to pool the agents' beliefs until a consensus has been reached. Recently, however, there has been a shift towards adopting a more pragmatic approach to achieve consensus formation in multi-agent systems, inspired by works in social epistemology whereby agents may obtain external information in the form of direct evidence from their environment as well as receiving information from other agents (Cho and Swami 2014; Douven 2019; Douven and Kelp 2011; Lee et al. 2018a).

A class of decision problems which has been widely studied in swarm robotics is the best-of-n problem whereby agents attempt to identify which is the best from a finite set of n possible options (Parker and Zhang 2009; Valentini et al. 2016, 2017). This is a popular example of a consensus formation problem (referred to as 'quorum sensing' in biological systems (List et al. 2009)) in which the agents seek to obtain evidence about each of the n options by exploring their environment. At the same time, the agents will encounter other agents and subject to constraints communicate their beliefs to one another in an attempt to reach an agreement about which option is the best. Over time, as agents repeat this process a system-wide consensus emerges and a clear decision naturally follows from having identified which is the best option.

Here we consider collective learning to be a more general distributed process in which agents attempt to learn the current state of their environment based on sparse evidence[1]. This evidence is obtained either directly from their environment, e.g. using their on-board sensing capabilities, or indirectly by fusing their beliefs with those of other agents. This fusion of beliefs between agents helps to propagate evidence through the system whilst also removing inconsistencies that might result from noisy or erroneous evidence. The fundamental distinction is that collective learning as a process does not presume a decision inherently follows from having reached a consensus about the environment state. For example, in the best-of-n problem a decision naturally follows from having reached a consensus about the best option due to the systematic discarding of information about the $n-1$ alternatives. By eliminating the other options as part of the learning process, the system can only act on the belief that the remaining option is the best. Indeed, many of the approaches to the best-of-n problem have so far restricted themselves to only considering the $n=2$ case. This limitation is often due to their having been inspired by, or even attempting to model directly, the solutions found in natural systems such as the nest site selection behaviours of social insects (Seeley and Buhrman 2001; Britton et al. 2002; Sumpter and Pratt 2009). Recently, however, we have seen increased interest in developing models that move us closer to the goal of deploying robot swarms for real-world applications, for example by proposing models for collective learning which consider more complex environments as

[1] Notice that the term 'learning' here is used to describe the process by which agents obtain evidence to inform their beliefs, or to *learn*, about their environment. This is not to be confused with similar terminology used in reinforcement learning or social learning.

well as their robustness to the presence of noise or error (Crosscombe et al. 2017; Lee et al. 2018b).

Search and rescue is a popular example of a well-suited application for swarm robotics as it requires the discovery of casualties in a sparse and often dangerous environment. The best-of-n problem fails to address this example application because it is only capable of considering the highest priority casualty as a result of the learning process. As mentioned above, information regarding the other casualties is most likely discarded and the decision-making process would have to be repeated for each remaining casualty. Alternatively, any deployed system ought to be able to accurately rank the casualties in order of severity in a timely manner. Ideally, the collective learning process should also be robust, e.g. to noise, error, or malfunction. This is because a common assumption in swarm robotics is that systems are composed of many cost-efficient robots that may suffer from reliability issues, but this can be counteracted by the deployment of a large number of robots to achieve robustness through redundancy (Schranz et al. 2020). Any proposed solution should therefore be robust to noisy or erroneous information and able to scale to large numbers of individuals.

In this paper, we propose a model of collective preference learning and apply it to the best-of-n problem. However, instead of only learning which option is the *best*, agents attempt to rank *all* of the n options by repeatedly obtaining evidence from the environment to form a preference ordering. While models for pairwise preference fusion have been proposed in the context of social networks (Brill et al. 2016; Hassanzadeh et al. 2013), these models employ a majority rule in order to reach consensus. In Sect. 2, we propose an iterative process of belief fusion in which pairs of agents combine their preference orderings by exploiting a third truth value to represent uncertainty. In Sect. 3, we demonstrate how over time the agents are able to learn a total ordering defined over $n = 10$ options, even when the environment is noisy. Then in Sect. 4, we discuss the implications of preserving the transitive closure of agents' beliefs during the learning process. The trade-off between speed and accuracy is an important one in applications such as search and rescue because the outcome of reaching a decision too slowly may be the same as ranking the priority of casualties incorrectly. As a result, we analyse the performance of our approach with respect to both accuracy and runtime as an indicator of the additional computational requirements of preserving transitivity. In Sect. 5, we study the limits of our approach when scaling to environments of different dimensions, i.e. from $n = 5, \dots, 25$ dimensions. Then in Sects. 6 and 7, we introduce two alternative models for collective preference learning under limited communications, e.g. when the bandwidth between agents is restricted due to hardware constraints. Specifically, Sect. 6 details a modified version of our original three-valued model adapted so that agents transmit only a subset of preference relations, while in Sect. 7 we introduce a probabilistic model whereby agents attempt to learn a probability distribution over the n options and deduce a ranking from that distribution. Finally, we provide some concluding remarks in Sect. 8.

2 A model for collective preference learning

Consider the basis of a best-of-n problem with n options $\mathcal{O} = \{o_1, \dots, o_n\}$. We assume that there is a true preference ordering on the options dependent on their relative quality and that this takes the form of a strict total order on $\mathcal{O}$, denoted by $\succ$. For example, in a search and rescue scenario there exists an inherent ordering of the casualties, ranking them from most to least in need of medical attention. Each agent possesses a belief

Table 1 The fusion operator $\odot$ applied to beliefs R and R'

	$R'_{i,j}$			
$R_{i,j}$	$\odot$	$\mathbf{0}$	$\frac{1}{2}$	$\mathbf{1}$
	$\mathbf{0}$	0	0	$\frac{1}{2}$
	$\frac{1}{2}$	0	$\frac{1}{2}$	1
	$\mathbf{1}$	$\frac{1}{2}$	1	1

about $\succ$ represented by an imprecise relation in the form of a $n \times n$ matrix R for which $R_{i,j} \in \{0, \frac{1}{2}, 1\}$. Here $R_{i,j} = 1$ means that the agent believes that the preference assertion $o_i \succ o_j$ is true, $R_{i,j} = 0$ means they believe it is false, and $R_{i,j} = \frac{1}{2}$ means that they are uncertain whether it is true or false. Since for each agent the relation R is intended to represent incomplete knowledge of the strict total order $\succ$, it would be natural to require R to satisfy transitivity, i.e. $R_{i,j} = 1$ & $R_{j,k} = 1$ implies $R_{i,k} = 1$. However, since we are assuming that evidence will always compare a particular pair of options, the transitive closure operation would then need to be performed after any such update to ensure that transitivity is preserved. This is computationally expensive and in Sects. 4, and 5, we investigate and compare collective preference learning both with and without the application of the transitive closure operation.

Belief fusion. We will consider pairwise fusion of agents' beliefs as follows: The binary operator $\odot$ is defined on $\{0, \frac{1}{2}, 1\}$ as given in Table 1. Now given R and R' corresponding to the beliefs of two agents, we define the fused belief matrix $R \odot R'$ in terms of the element-wise application of the $\odot$ operator so that $(R \odot R')_{i,j} = R_{i,j} \odot R'_{i,j}$. The fused belief $R \odot R'$ is then adopted by both agents forming a pairwise consensus.

Evidential updating. Evidence takes the form of a preference assertion $E =$ '$o_i \succ o_j$' upon receiving which an agent updates their belief matrix R to $R|E$ such that

$$(R|E)_{i,j} = 1,\ (R|E)_{j,i} = 0 \text{ and}$$
$$(R|E)_{u,v} = R_{u,v} \text{ for } (u,v) \neq (i,j),\ (u,v) \neq (j,i).$$

Updating in this manner does not take account of any additional information that can be inferred about $\succ$ from transitivity. Hence, ideally an agent who has just updated their belief should then apply the transitive closure operation. More specifically, the agent should take their new belief to be $(R|E)^+$ as obtained in this three-valued setting by making the minimum of changes to those elements of $R|E$ with truth value $\frac{1}{2}$ so as to ensure that the following holds: For all i, j,

- If $(R|E)^+_{i,j} = 1$ & $(R|E)^+_{j,k} = 1$ then $(R|E)^+_{i,k} = 1$
- If $(R|E)^+_{j,i} = 0$ & $(R|E)^+_{k,j} = 0$ then $(R|E)^+_{k,i} = 0$.

As noted above, there is a significant computational cost of applying this operator and we will investigate performance both with and without it.

In order to obtain evidence, agents must choose which options to investigate. We assume that agents select a pair of options $o_i, o_j \in \mathcal{O}$ to visit at random from those pairs about which they are uncertain, i.e. where $R_{i,j} = \frac{1}{2}$. Having selected two options to compare, agents will then either obtain some evidence E with probability r or learn nothing with probability $1 - r$, where r is an evidence rate quantifying the sparsity of evidence

in the environment. More formally, during each time step an agent with belief R shall update its belief such that

$$R = \begin{cases} R|E & : \text{ with probability } r \\ R & : \text{ with probability } 1 - r. \end{cases}$$

The process by which agents obtain evidence is also subject to noise; perhaps due to noisy sensors on-board the agents or merely a feature of the environment itself. We therefore introduce the notion of a comparison error as follows:

Definition 1 Comparison error The comparison error between any two options $o_i, o_j \in \mathcal{O}$ is given by

$$C_\lambda(o_i, o_j) = \frac{1}{2}\left(\frac{e^{-\lambda d} - e^{-\lambda}}{1 - e^{-\lambda}}\right),$$

where $d = \frac{|i-j|}{n}$ is the relative distance between options o_i and o_j in the true ordering $\succ$ and λ is a non-negative precision parameter in $\mathbb{R}^+$. In the case that $\lambda = 0$ then the comparison error becomes

$$\lim_{\lambda \to 0} C_\lambda(o_i, o_j) = \frac{1}{2}(1 - d).$$

Furthermore, as we increase λ then the comparison errors for all $o_i, o_j \in \mathcal{O}$ become

$$\lim_{\lambda \to +\infty} C_\lambda(o_i, o_j) = 0.$$

Notice that the comparison error depends upon the relative distance between the options in $\succ$. If the options being compared are closer together according to $\succ$, then they have a higher probability of being incorrectly compared while options which are further apart have a higher probability of being compared correctly. We scale the comparison errors by $\frac{1}{2}$ to ensure that we do not consider the case where the evidence obtained is more likely to be incorrect than correct. The comparison error therefore determines the probability with which the evidence takes the form of an incorrect preference assertion. That is, for $o_i, o_j \in \mathcal{O}$ where $o_i \succ o_j$:

$$E = \begin{cases} o_i \succ o_j & : \text{ with probability } 1 - C_\lambda(o_i, o_j) \\ o_j \succ o_i & : \text{ with probability } C_\lambda(o_i, o_j). \end{cases}$$

It may be useful to identify and compare the difficulty of learning $\succ$ for different values of λ as it may not be immediately obvious how this affects the level of precision, or inversely the level of noise, experienced by the agents. In Fig. 1, we plot C_λ against the distance between the options in $\succ$ for $n = 10$ options and different values of $\lambda = 1, \ldots, 100$. Here we see that $\lambda = 0$ is the case where the error decreases linearly as the distance between the options increases, while for larger values of lambda the comparison error decreases more rapidly as the distance between the options in $\succ$ increases. For $\lambda = 100$, there is no error when comparing even neighbouring options for $n = 10$. Furthermore, as the error between two options o_i, o_j varies depending both on λ and the distance between the options, $|i - j|$, according to $\succ$, it is useful to quantify the expected error in our model.

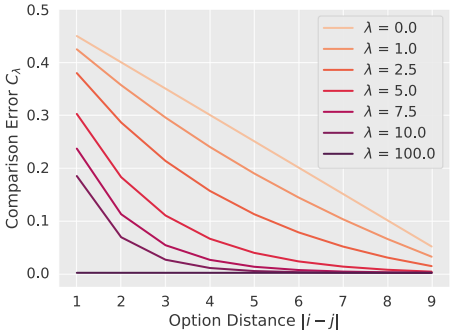


Fig. 1 Comparison errors across absolute distance between options $o_i, o_j \in \succ$ for different levels of precision $\lambda = 0, \ldots, 100$ and $n = 10$ options

Definition 2 Expected error The expected error of comparing any two options in $\mathcal{O}$ is the sum of all comparison errors produced by C_λ multiplied by their respective probabilities of occurring, as follows:

$$\mathbb{E}[C_\lambda] = \sum_{i=1}^{n-1} c_i p_i$$

where

$$p_i = 2\left(\frac{n-1}{n(n-1)}\right)$$

is the probability of any two options in $\mathcal{O}$ having a distance of i according to $\succ$, and

$$c_i = \frac{1}{2}\left(\frac{e^{-\frac{\lambda i}{n}} - e^{-\lambda}}{1 - e^{-\lambda}}\right)$$

is the comparison error for two options with distance i.

The expected error then provides us with a baseline for performance such that if the agents were to update based on evidence alone–that is, without any fusion taking place between the agents' beliefs–then the system would be expected to converge to an average error of $\mathbb{E}[C_\lambda]$, e.g. $\mathbb{E}[C_0] = 0.32$. If a population converges to an average error that is below the expected error, then we can say that the system's accuracy is being improved by the fusion process, while if the average error remains above the expected error then that system has performed worse than if agents were to conduct individual, rather than collective, learning.

To evaluate the performance of a population of agents attempting to learn the true preference ordering $\succ$, we propose to study the average error of the population as given in Definition 3. We assume that the true preference ordering $\succ$ can be represented by the $n \times n$ matrix R^* where $R^*_{i,j} \in \{0, 1\}$.

Definition 3 Average error The average error of a population of k agents is the normalised difference between each agent's belief R^a, for $a = 1, \ldots, k$, and the true preference ordering R^* averaged across the population as follows:

$$\frac{1}{k}\sum_{a=1}^{k}\frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j=i+1}^{n}\left|R_{i,j}^{a}-R_{i,j}^{*}\right|.$$

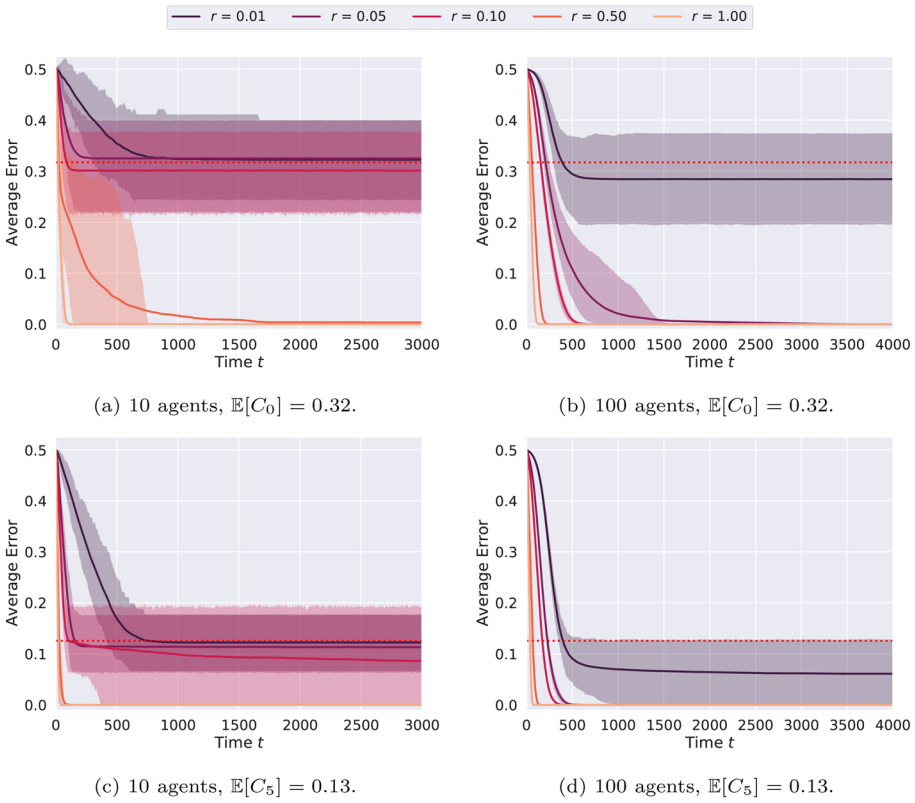
In the next section, we describe agent-based simulations of preference learning based on the three-valued model defined here.

3 Agent-based simulations

We consider an environment with n options, and for each experiment, we initialise a population of k agents where k varies between 10 and 100. Without loss of generality, we define the true ordering $\succ$ on $\mathcal{O}$ to be $o_1 \succ \cdots \succ o_n$. At initialisation, the agents begin in a state of complete ignorance about the true ordering $\succ$, i.e. at time $t = 0$ each agent holds the belief $R_{i,j} = \frac{1}{2}$ for all $i, j \in \{1, ..., n\}$. Notice that by Definition 3 such a belief, representing complete uncertainty, has an average error of 0.5 and therefore each population will begin with an average error of 0.5. Furthermore, if the population converges on the true ordering $\succ$ then the average error will be 0. While the rate at which agents combine their beliefs is fixed at one pair per time step, we define $r \in (0, 1]$ as the population-wide evidence rate which determines the frequency with which agents update based on direct evidence. That is, every agent has an equal probability r of successfully receiving evidence from the environment. Agents are also likely to experience noisy evidence with $\lambda \in [0, 100]$ denoting the extent to which the environment is noisy; notice that *larger* values of λ result in *higher* precision, or *less* noise during option comparisons. For a given set of parameter values, we run each experiment 100 times, averaging the results over those runs and each experiment runs for a maximum of 10,000 time steps or until the population converges. Variations across the 100 runs are shown as shaded regions representing the 10^{th} and 90^{th} percentiles. We define convergence as the beliefs of the population remaining unchanged for 100 interactions, where an interaction means either updating based on evidence or on the beliefs of other agents through application of the fusion operator defined in Table 1.

For our experiments, we adopt the 'well-stirred system' assumption from (Parker and Zhang 2009) which for the purpose of interaction translates to treating agents as nodes in a totally connected network. It is then possible for any agent to communicate with any other agent and each 'encounter' is modelled as an independent event. The reasoning behind this assumption is that in many swarm robotics applications, individual robots typically explore their environment stochastically based on local feedback received either directly from their environment or via other agents within communication range. Over time, as individuals continue to explore, the system will have mixed well enough as to be considered 'well-stirred' and the local neighbourhoods of each individual agent will have changed to account for the agents entering/exiting their range of communication. However, in this paper we do not take into account agents' spatial positions. Instead, we shall adopt the well-stirred system assumption going forward. Therefore, during any one time step we simply select an edge connecting two agents in the network before combining their beliefs using the fusion operator in Table 1 and this belief is adopted by both of the selected agents. From this assumption, it then follows that agent communication is not dependent upon which options each agent may have gathered evidence about prior to the belief fusion process, i.e. the agents remain well-stirred for the entire duration of the simulation.

In this section, we study the macro-level performance of our proposed approach to collective preference learning in the context of the best-of-n problem for $n = 10$. We present



(a) 10 agents, $\mathbb{E}[C_0] = 0.32$.

(b) 100 agents, $\mathbb{E}[C_0] = 0.32$.

(c) 10 agents, $\mathbb{E}[C_5] = 0.13$.

(d) 100 agents, $\mathbb{E}[C_5] = 0.13$.

Fig. 2 Average error over time for different evidence rates $r \in [0, 1]$. For $\lambda = 0$ and $\lambda = 5$, the expected errors $\mathbb{E}[C_0] = 0.32$ and $\mathbb{E}[C_5] = 0.13$ represent high and moderate noise scenarios, respectively. The red dotted line represents the expected error $\mathbb{E}[C_\lambda]$, and the shaded regions represent the 10^{th} and 90^{th} percentiles

both trajectory and steady-state results across a range of parameter settings before considering the effects of preserving transitivity of agents' beliefs in Sect. 4.

3.1 Convergence results for $n = 10$

Figure 2 shows the trajectories of the average error over time for populations of 10 and 100 agents. Each solid line represents a different evidence rate r between 0.01 and 1 while the red dotted line shows the expected error given λ according to Definition 2. In Fig. 2a, we present the dynamics for a high noise scenario with $\mathbb{E}[C_0] = 0.32$. For 10 agents and an evidence rate $r = 1.0$, the population converges to 0 average error in 123 time steps on average, while for $r = 0.5$ the average error slowly declines towards 0 and eventually converges after an average of 3,850 time steps. When the rate at which agents receive evidence is too low, e.g. $r \leq 0.5$, the population no longer converges to the true preference ordering and instead agents reach a steady state that is, on average, close to the expected error $\mathbb{E}[C_0] = 0.32$. Figure 2b shows the same trajectories but for 100 agents. Due to a larger population size, we see that the system benefits from increased exposure to evidence at the lower evidence rates. When $r = 0.01$, the population is still incapable of learning the

true preference ordering due to receiving an insufficient amount of evidence. However, as r increases the population learns the true preference in roughly 3,600 time steps for $r = 0.05$, around 930 time steps for $r = 0.1$ and at the higher end of the evidence range, with $r = 0.5$ and 1, the population requires just 325 and 162 time steps on average, respectively.

We observe in Fig. 2c that for a moderately noisy scenario with $\mathbb{E}[C_5] = 0.13$ and only 10 agents, an evidence rate of $r \geq 0.5$ is necessary for agents to learn the true preference ordering, requiring under 700 time steps and just 70 time steps on average for $r = 0.5$ and 1, respectively. For $r \leq 0.1$, the population performance is close to the expected error of 0.13. When the population is increased to 100 agents in Fig. 2d, the performance is similar for evidence rates $r \geq 0.05$ where the number of time steps required for the population to converge on the true preference ordering decreases as the evidence rate increases. For $r = 0.01$, where only one agent per time step on average successfully obtains evidence, the population fails to reliably converge to the true preference ordering.

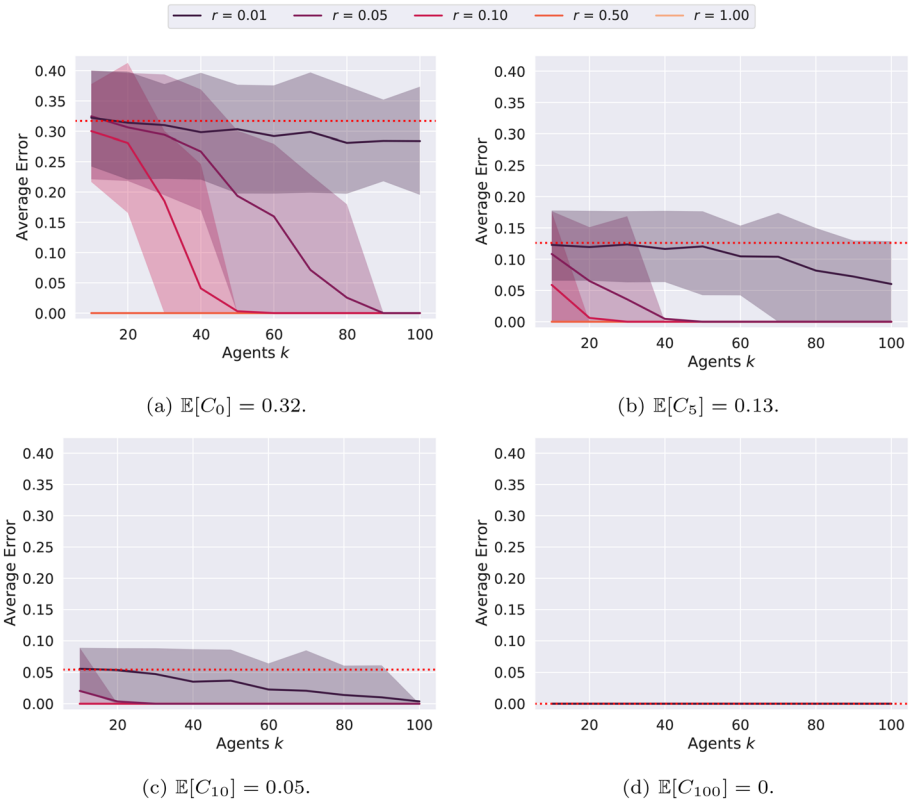
These trajectory results suggest that when the rate at which individual agents are able to obtain evidence is low, a larger population size is necessary to ensure that the system collectively obtains the evidence it needs to converge to an accurate ranking of the options. That is, the evidence rate r and the population size k are correlated in determining whether the system will successfully learn the true ordering of the options in $\mathcal{O}$.

Figure 3a to d shows the average error at steady state[2] for a decreasing amount of noise in the environment. As before, we vary the evidence rate r from 0.01 to 1. In Fig. 3a with $\mathbb{E}[C_0] = 0.32$ and $n = 10$ options, the probability of learning a false preference assertion when comparing two adjacent options, i.e. where $|i - j| = 1$ for $o_i, o_j \in \mathcal{O}$, is 0.45. We can see that for such a high level of noise and moderate evidence rates the population size plays an important role in governing convergence to the true preference ordering. When $r = 0.01$ for 10 agents, the population achieves an average error of 0.32 at steady state, equal to the expected error, and for 100 agents this is reduced to 0.28. At the other extreme when $r \geq 0.5$, the agents always converge to an average error of 0 for all population sizes. It is for evidence rates of $r = 0.05$ and 0.1 where we see that increasing the population size results in improved performance as the system is able to more reliably converge to 0 average error.

In Fig. 3b to d, we observe the expected behaviour where the average error decreases for all evidence rates as the environment becomes less noisy. In Fig. 3b with an expected error $\mathbb{E}[C_5] = 0.13$, all populations of $k = 10$ to 100 agents converge to 0 average error when the evidence rate $r \geq 0.5$. For $r = 0.1$, populations of 30 agents or more always converge to 0 average error, while for $r = 0.05$ a population of 50 agents or more is required. With an evidence rate $r = 0.01$, all population sizes produce an average error above 0 but below the expected error at steady state. In Figure 3c with $\mathbb{E}[C_{10}] = 0.05$, we see that for $r \geq 0.1$ all populations converge to the true preference ordering while for $r = 0.05$ populations of $k \geq 30$ agents also achieve 0 average error. Again, $r = 0.01$ proves ineffective at reliably learning the true preference ordering, instead achieving an average error of 0.004 with a population of 100 agents. Figure 3d shows that when the environment is essentially noise-free any population with 10 agents or more will learn the true preference ordering, converging to an average error of 0 for all evidence rates.

Generally, the model is robust to high levels of noise provided that the system is able to obtain evidence at a sufficient rate, either owing to a larger population or an environment in which evidence can be more frequently obtained. This is shown in Fig. 3a where large

[2] Either after the population has converged, or after 10,000 time steps.



(a) $\mathbb{E}[C_0] = 0.32$.

(b) $\mathbb{E}[C_5] = 0.13$.

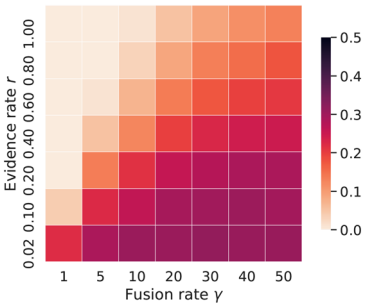
(c) $\mathbb{E}[C_{10}] = 0.05$.

(d) $\mathbb{E}[C_{100}] = 0$.

Fig. 3 Average error across different population sizes for different evidence rates $r \in [0, 1]$. The red dotted line represents the expected error $\mathbb{E}[C_\lambda]$ and the shaded regions represent the 10th and 90th percentiles

populations of 90 or more agents achieve a surprising level of accuracy despite the population only obtaining evidence at a rate of 5%. This is because in larger populations agents are more likely to encounter inconsistencies amongst their beliefs owing to having gathered conflicting pieces of evidence. By combining their beliefs according to Table 1, agents resolve these inconsistencies by becoming less certain about any pairwise relations that conflict before seeking additional evidence to determine their truth. Afterwards, agents will seek to obtain additional evidence and this slows the process of convergence, allowing the system to reconsider erroneous preference relations for as many times as is necessary until the population eventually reaches a global consensus. This is supported by Fig. 2 where populations tend to converge more slowly when the environment exhibits more noise.

For a complete picture of the dynamics of the model, we also show a heatmap of the fusion rate $\gamma = 1, \ldots, 50$ against evidence rate $r \in [0.02, 1]$ in Fig. 4. Here, the fusion rate γ determines the number of pairs of agents that fuse their beliefs to reach a pairwise consensus during each time step, e.g. $\gamma = 1$ means a single pair of agents fuse their beliefs each time step, while $\gamma = 50$ means that all 100 agents are paired up for belief fusion every time step. Until now, we have set $\gamma = 1$ as the standard fusion rate, but changed the population size which acted as a sort of proxy for the fusion rate. With a population size of 100, agents would fuse their beliefs far less regularly than in a population of 10 agents. In this figure,



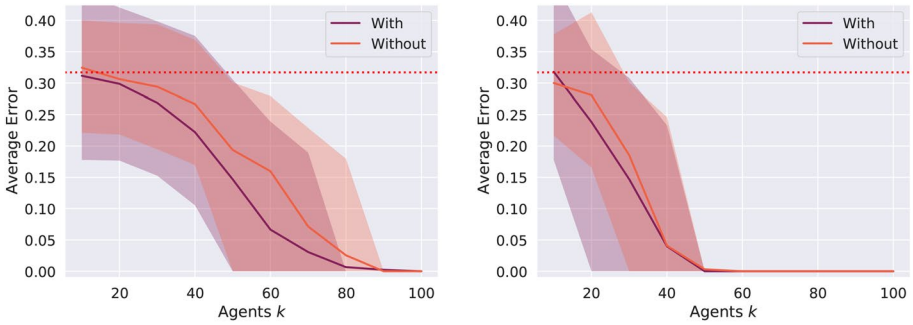
100 agents, $\mathbb{E}[C_0] = 0.32$.

Fig. 4 Average error compared across both fusion rate $\gamma = 1, \ldots, 50$ and evidence rate $r \in [0, 1]$. The fusion rate γ indicates the number of pairs of agents that are selected for belief fusion during each time step

we instead fix the population size at 100 agents and run the simulation experiments for different fusion rates and evidence rates. We see that, when agents are fusing their beliefs at a higher rate than the population can receive evidence, the accuracy of the system decreases to around the expected error $\mathbb{E}[C_0] = 0.32$. For the upper-left region of the heatmap, we see that having much higher evidence rates compared with the fusion rate results in convergence close to, or precisely to, the true ranking of $\mathcal{O}$. For example, when $\gamma = 1$ and only a single pair of agents fuse their beliefs per time step, an evidence rate of $r = 0.2$ is sufficient for convergence to the true ordering. When $\gamma = 5$, meaning that 10 agents are involved in the belief fusion process, an evidence rate of around $r = 0.8$ is suddenly required to prevent the increase in fusion from causing agents to converge prematurely to an inaccurate preference ordering. This also helps to explain why scaling the number of agents in the population helped to improve the accuracy of the model: when there is a larger number of agents in the system, there is an increase in the ratio of evidence gathering to belief fusion. The importance of evidence over the fusion process is generally true of most collective learning systems, as the fusion process' main purpose is to provide error correction and, in very sparse environments, the propagation of evidence in conjunction with error correction.

4 The effects of preserving transitivity

Until now, we have assumed that the combined processes of evidential updating and belief fusion between agents is sufficient to ensure that the population reaches a consensus about the true ordering $\succ$. While we have shown this to be true under certain conditions, including a sufficient evidence rate and population size, we have yet to address how our approach performs when we apply the transitive closure operator on agents' beliefs. Therefore, in this section we study the effects that preserving transitivity has on the population's ability to converge to the true preference ordering. Furthermore, we examine how the runtime performance is impacted by applying the transitive closure operation and therefore whether preserving transitivity of agents' beliefs is a worthwhile trade-off for real-world applications, particularly when decision-making is time-critical. Here we define runtime as the



(a) Evidence rate $r = 0.05$.

(b) Evidence rate $r = 0.1$.

Fig. 5 Comparisons of the average error across different population sizes with and without preservation of transitivity. We depict a high noise scenario with expected error $\mathbb{E}[C_0] = 0.32$, represented by the red dotted line. The shaded regions represent the 10th and 90th percentiles

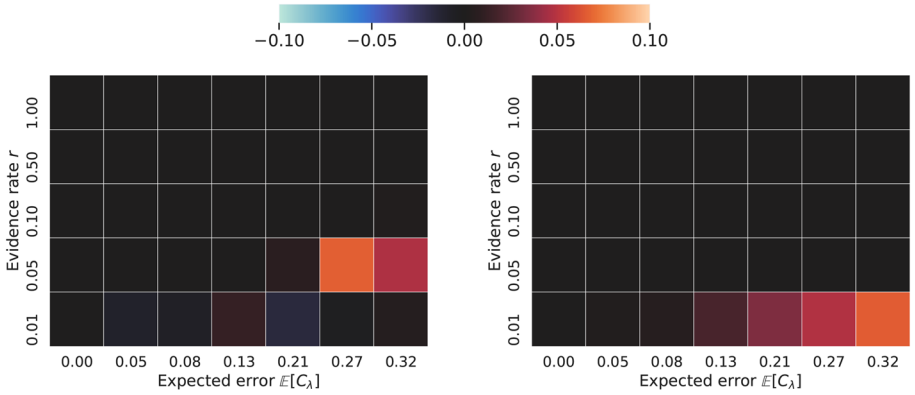
amount of CPU time in seconds that a process takes to complete one run of a simulation experiment[3].

4.1 Comparison results for applying the transitive closure operation

In Fig. 5, we present the average error over population size for a comparison between two distinct variants of our model: the purple line depicts performance while preserving transitivity of the agents' beliefs, while the orange line shows our model without applying the transitive closure operation. We have chosen to present a high noise scenario with $\mathbb{E}[C_0] = 0.32$ as this provides us with the largest range of separation between the two variants and allows us to observe more closely the difference in convergence to the true ordering. These results reflect the populations at the end of the simulation, either after 10,000 time steps or after the population has converged, whichever occurs first. We can see in Fig. 5a that preserving transitivity does indeed produce an noticeable performance gain when the rate at which agents receive evidence is low. For populations under 90 agents, i.e. for populations that do not converge to an average error of 0 at this evidence rate, we see that preserving transitivity leads to a reduction in average error. In some cases, such as for 60 agents, preserving transitivity can improve the average error by as much as 60% from 0.16 to 0.06. In Fig. 5b, we see that for a higher evidence rate $r = 0.1$ the impact on the accuracy of our approach is reduced. Applying the transitive closure operator results in slightly lower average error for smaller populations between 20 and 40 agents but the difference decreases with population size.

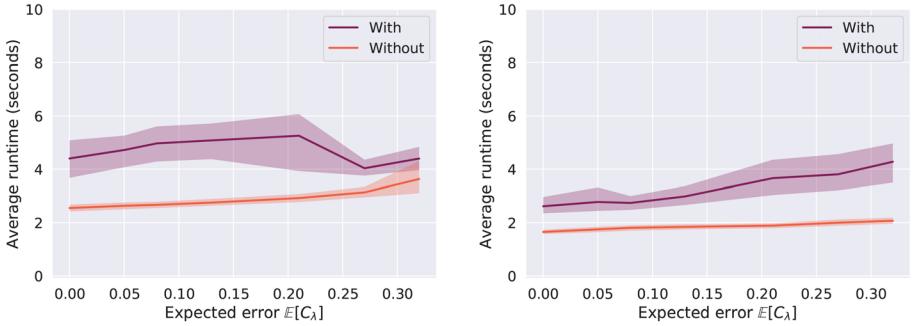
While it is useful to compare the two approaches for a particular combination of parameter settings, in Fig. 6 we show heatmaps of the difference in average error between applying and not applying the transitive closure operation, across increasing expected error $\mathbb{E}[C_\lambda]$ and evidence rates r. The warmer colours indicate the conditions for which preserving transitivity reduces the average error, while the cooler regions indicate an increase in

[3] The computer used to run these experiments is composed of an AMD Ryzen 9 3900X 12-core processor (3.8 GHz base clock, 4.6 GHz boost) with 32GB RAM (3600 Mhz).



(a) 50 agents. (b) 100 agents.

Fig. 6 Change in average error resulting from applying the transitive closure operation for $r \in [0.01, 1]$ and $\mathbb{E}[C_\lambda]$ with λ from 100 to 0. Warmer colours favour the preservation of transitivity, while cooler colours favour its omission



(a) Evidence rate $r = 0.05$. (b) Evidence rate $r = 0.1$.

Fig. 7 Average runtime in seconds against increasing expected error $\mathbb{E}[C_\lambda]$ for 100 agents, with and without the transitive closure. The shaded regions represent the 10th and 90th percentiles

average error when the transitive closure is performed. In Fig. 6a and b, we see that for higher expected error, corresponding to a noisier environment, and in conjunction with low evidence rates there is a clear benefit to applying the transitive closure operation. Particularly for high levels of noise, for a relatively low evidence rate $r = 0.05$ the transitive closure operation does reduce the average error. However, there are limited cases where this operation leads to noticeable performance improvements, and indeed for the lowest evidence rate $r = 0.01$ and a population of 50 agents there are small increases in average error when preserving transitivity, though this difference is minimal.

The above results suggest that the application of the transitive closure operation does improve performance for environments in which evidence is sparse and population sizes are small. Additionally, preserving transitivity is beneficial when the noise level is high but the rate at which agents obtain evidence remains low. However, it is equally important to understand the computational impact that a system is likely to incur as a result of applying

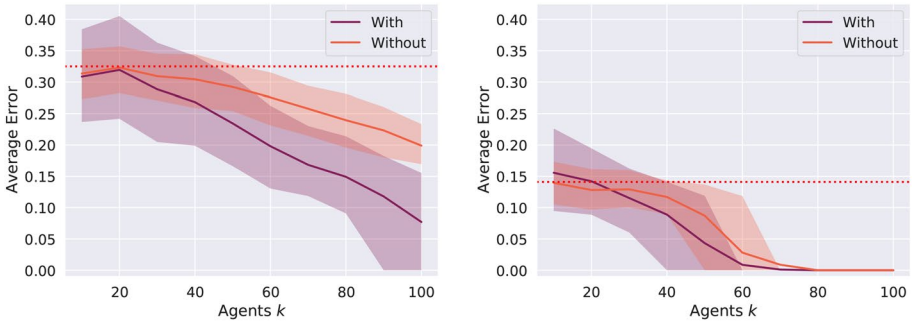
the transitive closure operation repeatedly during the collective learning process. To this end, we now discuss the runtime performance as an indication of the additional computational requirements of applying the transitive closure operation.

Figure 7 shows the average runtime in seconds against increasing expected error $\mathbb{E}[C_\lambda]$ for a population of 100 agents; chosen as the population converges in under the 10,000 time step limit, portraying consistent expected runtimes. We see that in Fig. 7a for $r = 0.05$ and expected error $\mathbb{E}[C_\lambda] = 0$ to 0.32, the resulting runtimes can be more than 2 seconds slower when preserving transitivity than when not. With this low evidence rate, we see that the runtimes both with and without transitivity are generally increasing with the noise level until $\mathbb{E}[C_\lambda] > 0.21$. Beyond this point, it seems that the two approaches behave similarly due to the combination of sparse and noisy evidence. In Fig. 7b for $r = 0.1$, we see a more monotonically increasing impact from applying the transitive closure operation. Without transitivity, the average runtime increases only marginally from 1.7 seconds to 2.1 seconds when the expected error is $\mathbb{E}[C_{100}] = 0$ and $\mathbb{E}[C_0] = 0.32$, respectively. When transitivity is preserved, the runtimes increase from 2.6 seconds to 4.3 seconds for the same expected errors. From a noise-free scenario to a high-noise scenario, the average runtime when applying the transitive closure operation grows to more than double that of our model without transitivity.

Broadly, we observe that preserving transitivity will result in increased runtimes but the degree to which slowdown occurs depends significantly upon the amount of noise present in the system. Noisy evidence increases the amount of conflict that occurs in a population as agents with inconsistent beliefs resolve these conflicts by adopting more imprecision in their beliefs. Doing so results in agents seeking additional evidence from the environment and each subsequent update of an agent's belief results in another transitive closure operation. When the system is too noisy and convergence to the true preference ordering does not occur, the system suffers from large increases in the runtime duration of the model. This might not pose a problem for one-off decisions where accuracy is paramount but in highly dynamic environments the collective learning process would need to be repeated often to capture changes in the environment. Although, when serialised, the presented average runtimes depict a large cost to applying the transitive closure operation, it should be noted that these results are only indicative of the additional computation that would be required in a distributed system where computation happens in parallel on each agent. Here, runtime acts as a proxy for computation and increased runtimes therefore depict a relative increase in computation required by each agent in the system. For swarm robotics in particular, each agent is typically equipped with only modest compute performance and a finite power source, presenting us with a trade-off between an increase in accuracy when learning the true preference ordering, and a decrease in performance due to additional computation being performed by each individual applying the transitive closure operation.

5 Scalability of collective preference learning

Many of the approaches to the best-of-n problem only consider the case of $n = 2$, the equivalent of a binary discrimination problem where agents must determine the truth or falsity of a single proposition (Kernbach et al. 2013; Parker and Zhang 2009, 2011;



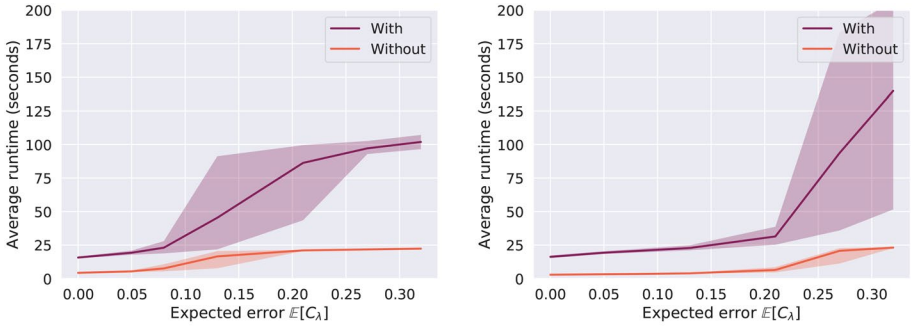
(a) $\mathbb{E}[C_0] = 0.32$. (b) $\mathbb{E}[C_5] = 0.13$.

Fig. 8 Comparisons of the average error for $n = 20$ options across different population sizes with and without preservation of transitivity and an evidence rate $r = 0.1$. The shaded regions represent the 10th and 90th percentiles

Prasetyo et al. 2019; Reina et al. 2015; Valentini et al. 2016, 2015)[4]. As the research community continues to make the case for applying swarm robotics to real-world problems, it is important to identify whether the proposed solutions can perform well in complex environments. Previously it has been shown that both the robustness and scalability of current approaches suffer in the presence of noisy and complex environments, for example in the weighted voter model (Crosscombe et al. 2017), with few models since having attempted to address these problems (Lawry et al. 2019; Lee et al. 2018b). So far we have shown that using our approach agents can learn an accurate ordering of $n = 10$ options and that the model is robust to high levels of noise under certain conditions, i.e. when the evidence rate $r \geq 0.1$ and the population consists of 50 or more agents. However, we must also consider the model's ability to scale to environments of increasing complexity. Due to our approach of ranking pairs of options in $\mathcal{O}$, agents have to learn $\frac{n(n-1)}{2}$ preference relations, or put another way, agents must obtain evidence about 45 pairwise relations until a system-wide consensus is reached. When we increase the complexity of our environment to 20 options, for example, agents must reach a consensus about 190 preference relations. In this section, we shall present results for $n = 5, \ldots, 25$ options and study the performance of our model with regard to both average error and runtime. We aim to highlight the extent of the scalability of this model while maintaining a focus on its robustness to noise in the environment.

Figure 8 shows the average error across population sizes from 10 to 100 agents, again comparing results with (purple) and without (orange) the preservation of transitivity. We study an evidence rate of $r = 0.1$ for two scenarios with high and low noise. We can see in Fig. 8a that for our high-noise scenario, with $\mathbb{E}[C_0] = 0.32$ and 20 options, an evidence rate of $r = 0.1$ is no longer sufficient to reach an average error of 0 for populations of 50 or more agents. In this scenario, the agents must gather evidence about twice as many options as for $n = 10$ and so lower evidence rates become insufficient to enable convergence to the true preference ordering $\succ$ when the population size is not adjusted. Of course, as the noise is reduced as in Fig. 8b, then for a moderately noisy scenario with $\mathbb{E}[C_5] = 0.13$ we begin

[4] Recent works considering $n > 2$ include (Crosscombe et al. 2017; Lawry et al. 2019; Lee et al. 2018a, b; Reina et al. 2017).



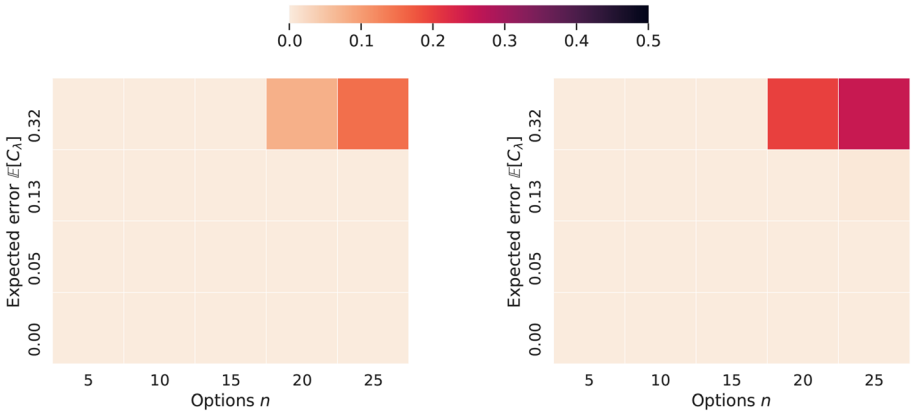
(a) $r = 0.05$. (b) $r = 0.1$.

Fig. 9 Average test runtime in seconds for $n = 20$ options and 100 agents across increasing expected error $\mathbb{E}[C_\lambda]$ with and without preservation of transitivity. The shaded regions represent the 10th and 90th percentiles

to see convergence to an average error of 0 for populations of 70 and above when transitivity is preserved, or 80 agents and above when we omit the transitive closure operation. The difference when applying the transitive closure operation is greater when the populations do not tend to converge, as is the case in Fig. 8a. When the environment is less noisy, this difference is reduced.

The difference between the two approaches is more apparent in Fig. 9 which shows the runtimes of our model with and without application of the transitive closure operation for $n = 20$ options and $k = 100$ agents. Here we show results with evidence rates $r = 0.05$ and 0.1 in Fig. 9a and b, respectively, plotted against expected error $\mathbb{E}[C_\lambda]$ from 0 to 0.32. Compared with the runtime results for 10 options, the difference in runtime between our approach with and without the preservation of transitivity is much greater. Specifically, without the preservation of transitivity and an evidence rate $r = 0.05$, runtimes range from 4.3 seconds for an expected error $\mathbb{E}[C_{100}] = 0$ to 22.3 seconds for $\mathbb{E}[C_0] = 0.32$. With transitivity, the average runtimes are dramatically increased, ranging from 15.7 seconds to 101.9 seconds for the aforementioned expected errors, respectively. Clearly, with $n = 20$ options the additional time required to compute the transitive closure of agents' beliefs is strongly increasing with the extent to which the evidence is noisy. Similar results can be seen in Fig. 9b for an evidence rate $r = 0.1$. This is due to the much larger number of preference relations that must be considered by the transitive closure operation each time a pair of agents fuse their beliefs. For example, with 10 options there were 45 preference relations to consider, whereas here with 20 options there are now 190 preference relations that must be considered. These extended runtimes are the result of additional computation that must be performed to calculate the transitive closure of an agent's fused belief and is therefore indicative of the additional computation that each agent would need to perform. For noisy environments, the additional computation required to form the transitive closure therefore increases exponentially with the number of options being considered.

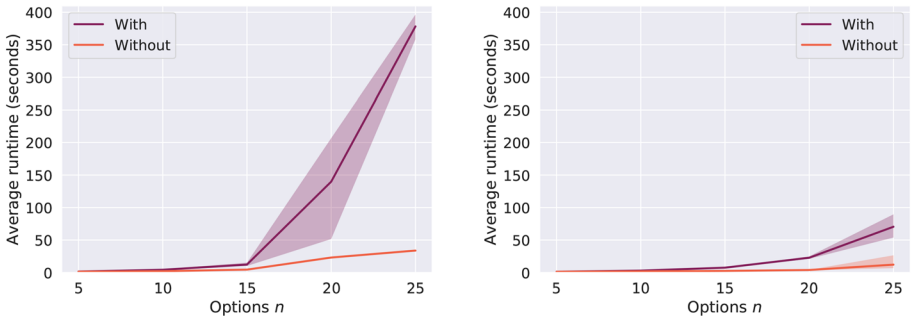
It is clear that there is an appreciable reduction in the average error when transitivity of agents' beliefs is preserved but this comes at the cost of large increases in runtime. In the extreme case, for an evidence rate $r = 0.1$, $\mathbb{E}[C_0] = 0.32$ and a population of 100 agents you can decrease the average error of the population's learning process by 53% from 0.19 to 0.08 at the cost of a runtime increase of 608% on average, going from 23 seconds to 140



(a) With transitive closure.

(b) Without transitive closure.

Fig. 10 Average error compared against number of options n and expected error $\mathbb{E}[C_\lambda]$ for λ from 100 to 0. Results are for a population of 100 agents with $r = 0.1$



(a) $r = 0.1$ and $\mathbb{E}[C_0] = 0.32$.

(b) $r = 0.1$ and $\mathbb{E}[C_5] = 0.13$.

Fig. 11 Average test runtime in seconds across the number of options n for 100 agents and different expected error $\mathbb{E}[C_\lambda]$. The shaded regions represent the 10th and 90th percentiles

seconds. However, provided enough agents can be deployed and the environment is not too sparse for agents to obtain evidence at a suitable rate, both approaches will converge to an average error of 0 while our initial model without transitivity will learn the true ordering in less time.

To provide a broader overview of the scalability of the proposed model for collective learning, we also include Fig. 10 where, for a population of 100 agents and an evidence rate $r = 0.1$, we show the average error across both the number of options n and the expected error $\mathbb{E}[C_\lambda]$ according to the noise level λ. Clearly, as the number of options increases such that $n \geq 20$ we begin to see the populations failing to consistently learn the true preference ordering $\succ$ for a high expected error $\mathbb{E}[C_0] = 0.32$. In Fig. 10a and b, we see that the application of the transitive closure operation does lead to a reduction in the average error in the cases where the populations fail to reliably reach a consensus. We see in Fig. 11a, however, that for the same expected error $\mathbb{E}[C_0] = 0.32$ the preservation of transitivity is

extremely costly in terms of runtime performance. Figure 11 shows the average runtime in seconds for $n = 5, \ldots, 25$ options for both $\mathbb{E}[C_0] = 0.32$ and $\mathbb{E}[C_5] = 0.13$. On the other hand, Fig. 11a and b shows that a reduction in noise produces a significant reduction in the average runtime when preserving transitivity. The scalability of our approach therefore depends on the avoidance of applying the transitive closure operation during the simulation when the level of noise is very high. If the evidence in the environment is subject to lower levels of noise, e.g. for $\mathbb{E}[C_5] = 0.13$, then this runtime cost is significantly reduced but for the decrease in average error it is difficult to make a case for the preservation of transitivity in this model.

6 Robustness to limited communications

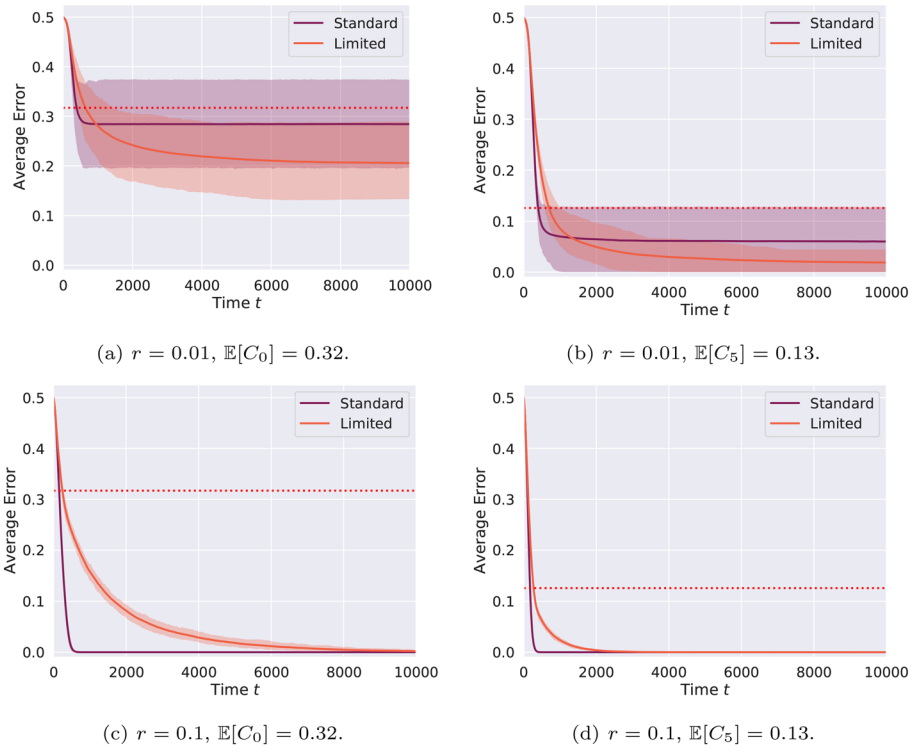
An important aspect of systems designed for collective learning is their utilisation of local communications to achieve system-level behaviours. Though many robotic platforms at present already possess reasonably sufficient communications capabilities, the bandwidth between individuals is not limitless. For example, Kilobots can communicate with other Kilobots within a radius of up to 10 cm at a rate of up to 30 kb/s (Rubenstein et al. 2014)[5]. Bandwidth limitations between agents therefore pose a potential hurdle for multi-agent systems where frequent communication is crucial to their function. In Sect. 2, we proposed a model of collective preference learning that relies upon pairs of agents sharing their full beliefs, composed of up to $\frac{n(n-1)}{2}$ preference relations, during a single time step so as to fuse their beliefs. While it would be possible to have agents split their preference set into multiple messages to be aggregated by the receiving agents, this kind of solution relies upon platform-specific implementations. Instead, in this section we propose a modification of our initial model that limits agents to communicating only a subset of their belief, specifically a belief with only n preference relations and demonstrating that while the convergence times are impacted by such an approach, the performance of the model in terms of accuracy at least meets, if not exceeds, that of our initial model.

As before, each agent possesses a belief about the true preference ordering $\succ$ represented by an $n \times n$ matrix R with $R_{i,j} \in \{0, \frac{1}{2}, 1\}$. To achieve a bandwidth-limited version of our proposed model, we make the following alteration to the belief fusion process: Until now, two agents would form a pairwise consensus by sharing their full beliefs R, R' with one another and both agents would then adopt the fused belief $R \odot R'$. Instead, we propose that each agent selects up to n preference relations that they believe to be true, at random, and share those with the other agent. Consider the set of clear (precise) preferences in R given by $C = \{(i,j) : j > i, R_{i,j} \neq \frac{1}{2}\}$. If $|C| > n$ then the agent selects $P \subset C$ such that $|P| = n$, otherwise they select $P = C$. The agent then forms $\tilde{R}$ where

$$\tilde{R}_{i,j} = \begin{cases} R_{i,j} & : (i,j) \in P \\ \frac{1}{2} & : \textit{otherwise}. \end{cases}$$

The agent transmits $\tilde{R}$ and at the same time receives $\tilde{R}'$, generated from R' based on the same process described above. Notice that transmitting $\tilde{R}$ only requires transmitting $|P| \leq n$

[5] This is the stated maximum communication rate of the Kilobots. However, due to practical limitations, the actual communication rate is often lower. For example, with up to 9 bytes per message, if the Kilobots update at a rate of 2 time steps per second, then the actual communication rate will be 144 b/s.



(a) $r = 0.01$, $\mathbb{E}[C_0] = 0.32$.

(b) $r = 0.01$, $\mathbb{E}[C_5] = 0.13$.

(c) $r = 0.1$, $\mathbb{E}[C_0] = 0.32$.

(d) $r = 0.1$, $\mathbb{E}[C_5] = 0.13$.

Fig. 12 Average error compared between the standard model (purple) and the bandwidth-limited model (orange) against time t for $n = 10$ options and 100 agents. The red dotted line represents the expected error $\mathbb{E}[C_\lambda]$ and the shaded regions represent the 10th and 90th percentiles

distinct preferences, i.e. $R_{i,j}$ for $(i,j) \in P$, since the receiving agent can adopt the protocol of assuming the value of $\frac{1}{2}$ for any unspecified preferences. Finally, the agents update their beliefs such that the agent with belief R adopts $R \odot \tilde{R}'$ as its new belief, while the other agent adopts $\tilde{R} \odot R'$ as its new belief. Unlike in our original model in which both agents adopted the same fused belief $R \odot R'$, the agents are instead adopting beliefs based on having received only partial preference orderings held by the other agents, which are likely to differ from one another. This means that agents no longer adopt a pairwise consensus as a result of the belief fusion process.

6.1 Comparing the standard and bandwidth-limited models

Figure 12 shows the average error across time for 10 options. Here we compare the standard model with the bandwidth-limited model for different evidence rates r and expected errors $\mathbb{E}[C_\lambda]$. We notice immediately from Fig. 12a that in a highly noisy environment with very sparse evidence, i.e. $\mathbb{E}[C_0] = 0.32$ and $r = 0.01$ the bandwidth-limited model, whereby agents may only communicate up to n preference relations, actually outperforms the standard model. As is clear from the figure, the limited model is much slower to converge but does reduce error to below that of the standard model, where the standard model reaches an average error of 0.28 while the limited model reaches an average of 0.2 at steady state. This

is due to agents only communicating a subset of their beliefs, which prevents agents from reaching a pairwise consensus during the fusion process and slowing convergence. These slower convergence times allow for agents to reach more accurate decisions by ensuring that, at such a low evidence rate, the fusion process does not prematurely lead to the population adopting a more erroneous preference ordering. These dynamics are very similar for the same low evidence rate $r = 0.01$ but reduced noise in Fig. 12b. That is, the limited model eventually outperforms the standard model, which converges faster but to a less accurate preference ordering. For a higher evidence rate of $r = 0.1$, as shown in Fig. 12c and d, the two models are equal in their accuracy, but with high noise the limited model converges much more slowly than in a less noisy scenario.

An advantage of our bandwidth-limited approach is that the subset of preference relations that form the communicated belief $\tilde{R}$ could be chosen in a non-random manner to maximise the number of preference relations that would be added under the transitive closure. Put another way, let's suppose an agent possesses a precise belief such that the agent believes it has identified a strict total ordering over all n options, i.e. $R_{i,j} \neq \frac{1}{2}$ for all i, j where $i \neq j$. Then the agent would only need to communicate $n - 1$ preference relations for the receiving agent to be able to generate the full set of preference relations and reform R. Additionally, in this model the amount of information being sent between agents can be tuned to adjust the speed vs. accuracy trade-off. For example, provided that the slower convergence times can be tolerated, the communication bandwidth could be purposefully restricted to enable slower, more accurate collective learning when preferred, while unlimited communication could be enabled should the scenario require faster, less accurate collective learning. The approach taken would therefore be specific to the context in which the system is deployed.

7 Probabilistic preference learning for limited communications

Following on from Sect. 6 in which we modified our original model to function under limited communications, with this same aim in mind we now consider an alternative probabilistic model whereby agents attempt to learn a probability distribution over the n options and therefore need only transmit n real values during the belief fusion process. That is, each agent holds a belief $B : \mathcal{O} \to [0, 1]^n$ where $B(o_i)$ represents the probability with which the agent believes option o_i to be the best. Of course, one immediate problem encountered when using a quantitative representation is the need to assign numerical values to each of the n options. Given that we are interested in ranking the options we first assign a quality value to each option such that for each $o_i \in \mathcal{O}$ the associated quality $q_i = \frac{i}{n+1}$. This way, the option qualities are ranked according to $\succ$ such that $q_n > \cdots > q_1$.

Another problem arising from this representation is the need for a fusion operator which preserves the order of the options based on their respective qualities. While a probabilistic model of belief may be well suited for learning which option is the best (Lee et al. 2018b), maintaining a preference ordering over $\mathcal{O}$ has to our knowledge not yet been considered in this context. In this probabilistic setting, we combine agents' beliefs using the probability fusion function in Definition 4.

Definition 4 Probability fusion function Let $\mathbb{P}$ denote the set of all probability distributions on $\mathcal{O}$. Then a probability fusion function is a function $c : \mathbb{P}^2 \to \mathbb{P}$. In particular, we will focus on the product fusion function given by: for $o_i \in \mathcal{O}$,

Table 2 The expected errors $\mathbb{E}[C_\lambda]$ and associated standard deviation σ for different noise values λ

λ	**0**	**1**	**2.5**	**5**	**7.5**	**10**	**100**
$\mathbb{E}[\boldsymbol{C}_\lambda]$	0.32	0.27	0.21	0.13	0.08	0.05	0.0
$\boldsymbol{\sigma}$	0.47	0.33	0.21	0.12	0.09	0.07	0.0

$$c(B_1, B_2)(o_i) = \frac{B_1(o_i)B_2(o_i)}{\sum_{j=1}^{n} B_1(o_j)B_2(o_j)}.$$

This established operator (Bordley 1982) has recently been studied by Lee et al. (2018a) in the context of the best-of-n problem. However, in order to prevent agents reaching absolute certainty about any particular option, i.e. probabilities of either 0 or 1, we apply a small weighting of the agent's belief towards ignorance as represented by the uniform distribution over $\mathcal{O}$. This helps the agents to maintain a preference ordering. More formally, immediately following any application of the probabilistic fusion function, the agents update their beliefs as follows: for $o_i \in \mathcal{O}$,

$$B(o_i) = \alpha\frac{1}{n} + (1-\alpha)B(o_i)$$

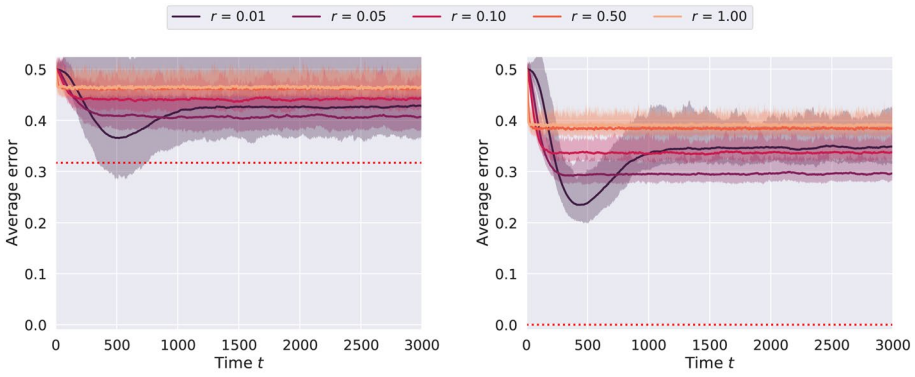
where $\alpha \in [0, 0.5]$ is a dampening term typically set to a small value, i.e. $\alpha = 0.1$ as in our simulation experiments.

In this model, agents obtain evidence from the environment in the following form:

$$B_E(o_j) = \begin{cases} \frac{1-q_i-\epsilon}{n} & : j \neq i \\ \frac{(n-1)(q_i+\epsilon)+1}{n} & : j = i \end{cases}$$

where ϵ is a normally distributed random variable with mean 0 and standard deviation σ, and B_E is the linear combination of the probability distribution with $B(o_i) = 1$ and the uniform distribution with weights $q_i + \epsilon$ and $1 - (q_i + \epsilon)$, respectively. Notice that under this noise model the error is applied to each option that is investigated as opposed to each *pair* of options as in our initial model. Then, the agent combines its current belief with the evidence using the probabilistic fusion operator, adopting $c(B, B_E)$ as its new belief before dampening its belief as above. Unlike in the original model in which agents only seek to obtain evidence when $R_{i,j} = \frac{1}{2}$, there is no natural way to determine whether agents should or should not seek additional evidence. Therefore, agents seek evidence about one option $o_i \in \mathcal{O}$ at random according to the evidence rate r. This is a clear weakness of this probabilistic approach as the agents will continue to seek additional evidence without intervention.

At time $t = 0$, agents have total ignorance about the options represented by their beliefs being the uniform distribution, i.e. $B(o_i) = \frac{1}{n}$ for $i = 1, \ldots, n$. Just as we described in our initial model from Sect. 2, agents then begin a process of gathering evidence and fusing their beliefs with other agents using the probabilistic processes described above. To closely align the probabilistic model with our original model, the noise value ϵ is drawn from a normal distribution with σ chosen to reflect the associated expected error $\mathbb{E}[C_\lambda]$. In Table 2, we show the values of σ which most closely match the expected errors for different precision values λ. To determine the average error in this model, we simply extract the preference ordering from the agent's belief B, where an agent believes $o_i \succ o_j$ if $B(o_i) > B(o_j)$ and believes $o_j \succ o_i$ if $B(o_i) < B(o_j)$. In the case that $B(o_i) = B(o_j)$ the agent has no preference between options o_i and o_j, e.g. as when $R_{i,j} = \frac{1}{2}$ in our original model.



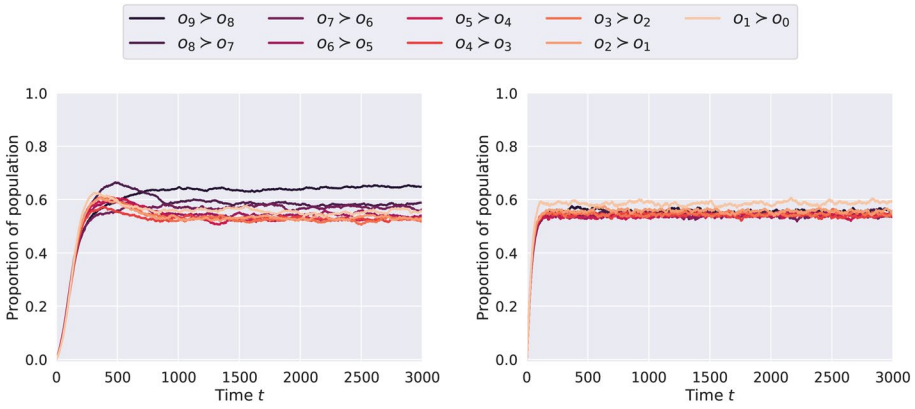
(a) $\sigma = 0.47$, $\mathbb{E}[C_0] = 0.32$. (b) $\sigma = 0$, $\mathbb{E}[C_{100}] = 0$.

Fig. 13 Average error over time for 100 agents and different evidence rates $r \in [0, 1]$. For $\sigma = 0.47$ and $\sigma = 0$, the expected errors $\mathbb{E}[C_0] = 0.32$ and $\mathbb{E}[C_{100}] = 0$ represent high and zero noise scenarios, respectively. The red dotted line represents the expected error $\mathbb{E}[C_\lambda]$ and the shaded regions represent the 10^{th} and 90^{th} percentiles

7.1 Convergence results with probabilistic beliefs

We now present summary results for the probabilistic model for collective learning. In Fig. 13, we show the average error of the population over time for different evidence rates r, with shaded regions depicting the 10^{th} and 90^{th} percentiles. Specifically, Fig. 13a shows the average error for a high-noise scenario where $\mathbb{E}[C_0] = 0.32$. Under high noise, the probabilistic model struggles to converge to an average error below 0.4 for all evidence rates. For $r = 0.01$, we see that the average error initially decreases to below 0.4 at around 500 time steps, before increasing again to converge at a higher average error than for $r = 0.05$. This is most likely owed to the system continuing to obtain evidence due to the agents lacking a mechanism for determining their certainty. When we move to a noise-free scenario as shown in Fig. 13b, we do see an improvement in average error across the different evidence rates r, with each showing similar dynamics as before, e.g. for $r = 0.01$, there is a large decrease in average error before 500 time steps before the average error increases again, this time exceeding even $r = 0.1$ on average. Given that the evidence is always accurate in this scenario, it is clear that this model is ineffective for consistently learning the true preference ordering.

An additional perspective of the convergence dynamics of the probabilistic model is presented in Fig. 14. In this figure, we show the proportion of the population with clear preferences $o_i \succ o_j$ based on the agents believing $B(o_i) > B(o_j)$ for $o_i, o_j \in \mathcal{O}$ and where $i > j$. That is, higher probability is assigned to option o_i than option o_j, indicating a clear preference in o_i over o_j. We can conclude from these dynamics that the probabilistic model is not suited to maintain a precise and accurate preference ordering over the $n = 10$ options, even for the noise-free scenario shown here. For the lowest evidence rate $r = 0.01$ shown in Fig. 14a, we see that the population does better at accurately learning the ordering between the options with higher quality, e.g. $o_9 \succ o_8$, than it does for options of lower quality. However, as we increase the evidence rate to $r = 0.1$ the performance for all options narrows to be roughly equal. Notice that we have chosen to



(a) $r = 0.01$. (b) $r = 0.1$.

Fig. 14 Proportion of the population which believes $B(o_i) > B(o_j)$, indicating preferences $o_i \succ o_j$ for all options $o_i, o_j \in \mathcal{O}$ where $i < j$, against time. We show a noise-free scenario for a population of 100 agents where $\sigma = 0$ and $\mathcal{E}[C_{100}] = 0$

omit the error bars here due to the extremely large variation between runs and generally poor accuracy performance of the probabilistic model.

In this section, we have proposed a probabilistic model based on a well-established fusion operator and modified the convergence behaviour using a dampening term to preserve an ordering over the n options. The results show that this probabilistic operator is not well suited to a collective preference learning problem due to its inability to maintain an accurate preference ordering over the options. Preliminary studies were also conducted using a simple averaging operator which appeared to perform similarly to the presented model. However, there may be other fusion operators that work well in this setting which we have not considered here.

8 Conclusions and future work

We have proposed a model for collective preference learning in the context of the best-of-n problem and demonstrated the ability of a population of agents to rank *all* of the n options during the learning process. Our model leads to highly robust consensus formation in populations of 10 to 100 agents and is scalable up to $n = 20$ options before the accuracy of the model is affected by high levels of noise in the evidential updating process. When evidence is sparse and the level of noise is high, the robustness of our model is improved by performing the transitive closure operation after agents fuse their beliefs. However, this operation is computationally expensive and the transitive closure is not strictly required to learn the true preference ordering, thereby offering a trade of increased convergence times and computational power for increased accuracy. This approach to collective learning may therefore be well suited to applications in noisy environments or for systems where large numbers of cost-effective but noisy sensors are deployed to identify the true state of the world. Using our approach, the system can accurately reach a consensus about the state of its environment despite high noise and/or a low frequency of obtaining evidence. We also demonstrated that our approach can be

adapted to perform well when the communication bandwidth between agents is limited, provided that longer convergence times are tolerable for the given application.

Furthermore, we have presented an alternative model in which agents' beliefs are represented by probability distributions and demonstrated that this approach performs poorly when agents must collectively rank the options, despite having previously been shown to perform well on the best-of-n problem and also only requiring agents transmit n real values as opposed to $\frac{n(n-1)}{2}$ pairwise preferences. In the context of ranking all n options, our proposed three-valued model performs well for a variety of scenarios and additional modifications can be made to better fit a range of conditions, such as sparse environments or hardware platforms that impose limitations on agents' communications.

Scalability is only one aspect to consider for deploying systems into complex environments: adaptability is another. More recently, Prasetyo et al. (2019) have explored the consensus formation problem in dynamic environments where one-off decisions are no longer sufficient and repeating the whole collective learning process when one feature of the environment changes is wasteful. Our proposed model could be easily extended by adopting a concept of "preference decay" to ensure that old information is periodically discarded and new evidence is gathered to maintain accuracy in the event that the environment changes.

Following this work, we plan to implement our model in a physical simulator. The added spatial element would provide agents with additional constraints such as ranges of communication and observation which shall affect the ability of agents to communicate with the rest of the system and to obtain evidence about specific locations of interest. Defining a range of communication would inform the underlying network topology over which agents would be required to communicate for the purpose of belief fusion. An observation range would act to have agents incur a cost when travelling between locations in order to obtain evidence about the environment. These aspects alone would further inform the suitability of our model to the collective learning problem, as well as allow us to better compare different approaches that exist in more abstract forms but have yet to prove their applicability to collective learning applications. Recently, Crosscombe and Lawry (2021) have shown that systems modelled with total connectivity behave sub-optimally on a similar collective learning task, demonstrating instead that limited connectivity leads to improved performance in this context. This is evidence that this simplifying assumption fails to adequately capture the dynamics of swarm systems. Therefore, through physical simulations we intend to investigate general principles for constructing effective communication network topologies for swarm systems, including the investigation of network topologies determined by the range of communication or even the pre-allocation of neighbourhoods with varying degrees of connectivity, regularity, etc.

Through physical simulations, we will also investigate the performance of our model when there is an inherent cost associated with agents exploring their environment. At present, the connectivity of agents is not dependent upon their range of communication and therefore there is no cost associated with agents exploring distant options that may place them out of range of other agents. As agents must travel between the pair of options in order to obtain evidence about them, some pieces of evidence will necessarily take longer to obtain and therefore be associated with a higher cost. We intend to explore how this impacts the performance of our model, particularly with regard to the preservation of transitivity which may provide our agents with a means of avoiding costly option comparisons.

Acknowledgements This work was funded and delivered in partnership between Thales Group, University of Bristol and with the support of the UK Engineering and Physical Sciences Research Council, ref.

EP/R004757/1 entitled “Thales-Bristol Partnership in Hybrid Autonomous Systems Engineering (T-B PHASE)”.

References

Baronchelli, A. (2018). The emergence of consensus: A primer. *Royal Society Open Science, 5*(2), 172189. https://doi.org/10.1098/rsos.172189.

Bordley, R. F. (1982). A multiplicative formula for aggregating probability assessments. *Management Science, 28*(10), 1137–1148. https://doi.org/10.1287/mnsc.28.10.1137.

Brambilla, M., Ferrante, E., Birattari, M., & Dorigo, M. (2013). Swarm robotics: A review from the swarm engineering perspective. *Swarm Intelligence, 7*(1), 1–41.

Brill, M., Elkind, E., Endriss, U., & Grandi, U. (2016). Pairwise diffusion of preference rankings in social networks. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI 2016)* (pp. 130–136).

Britton, N. F., Franks, N. R., Pratt, S. C., & Seeley, T. D. (2002). Deciding on a new home: How do honeybees agree? In *Proceedings: Biological Sciences* (Vol. 269(1498), pp. 1383–1388).

Cho, J., & Swami, A. (2014). Dynamics of uncertain opinions in social networks. In *2014 IEEE Military Communications Conference* (pp. 1627–1632).

Crosscombe, M., & Lawry, J. (2021). The impact of network connectivity on collective learning. In *Proceedings of the 15th International Symposium on Distributed Autonomous Robotic Systems (DARS)*. Cham: Springer International Publishing.

Crosscombe, M., Lawry, J., Hauert, S., & Homer, M. (2017). Robust distributed decision-making in robot swarms: Exploiting a third truth state. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 4326–4332).

DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association*, *69*(345), 118–121.

Douven, I. (2019). Optimizing group learning: An evolutionary computing approach. *Artificial Intelligence, 275,* 235–251. https://doi.org/10.1016/j.artint.2019.06.002.

Douven, I., & Kelp, C. (2011). Truth approximation, social epistemology, and opinion dynamics. *Erkenntnis, 75*(2), 271. https://doi.org/10.1007/s10670-011-9295-x.

Hassanzadeh, F. F., Yaakobi, E., Touri, B., Milenkovic, O., & Bruck, J. (2013). Building consensus via iterative voting. In *2013 IEEE International Symposium on Information Theory* (pp. 1082–1086).

Kernbach, S., Häbe, D., Kernbach, O., Thenius, R., Radspieler, G., Kimura, T., & Schmickl, T. (2013). Adaptive collective decision-making in limited robot swarms without communication. *The International Journal of Robotics Research, 32*(1), 35–55. https://doi.org/10.1177/0278364912468636.

Lawry, J., Crosscombe, M., & Harvey, D. (2019). Epistemic sets applied to best-of-n problems. In G. Kern-Isberner & Z. Ognjanović (Eds.), *Symbolic and Quantitative Approaches to Reasoning with Uncertainty* (pp. 301–312). Cham: Springer International Publishing.

Lee, C., Lawry, J., & Winfield, A. (2018a). Combining opinion pooling and evidential updating for multi-agent consensus. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-2018), International Joint Conferences on Artificial Intelligence Organization* (pp. 347–353), https://doi.org/10.24963/ijcai.2018/48.

Lee, C., Lawry, J., & Winfield, A. (2018b). Negative updating combined with opinion pooling in the best-of-n problem in swarm robotics. In M. Dorigo, M. Birattari, C. Blum, A. L. Christensen, A. Reina, & V. Trianni (Eds.), *Swarm Intelligence*, LNCS (Vol. 11172, pp. 97–108). Cham: Springer International Publishing.

Lehrer, K., & Wagner, C. (1981). Rational consensus in science and society: A philosophical and mathematical study. Pallas paperback, Springer Netherlands.

List, C., Elsholtz, C., & Seeley, T. D. (2009). Independence and interdependence in collective decision making: An agent-based model of nest-site choice by honeybee swarms. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1518), 755–762.

Parker, C. A. C., & Zhang, H. (2009). Cooperative decision-making in decentralized multiple-robot systems: The best-of-N problem. *IEEE/ASME Transactions on Mechatronics, 14*(2), 240–251. https://doi.org/10.1109/TMECH.2009.2014370.

Parker, C. A. C., & Zhang, H. (2011). Biologically inspired collective comparisons by robotic swarms. *The International Journal of Robotics Research, 30*(5), 524–535. https://doi.org/10.1177/0278364910397621.

Perron, E., Vasudevan, D., & Vojnović, M. (2009). Using three states for binary consensus on complete graphs. In *Proceedings—IEEE INFOCOM* (pp. 2527–2535), https://doi.org/10.1109/INFCOM.2009.5062181.

Prasetyo, J., De Masi, G., & Ferrante, E. (2019). Collective decision making in dynamic environments. *Swarm Intelligence, 13*(3–4), 217–243.

Reina, A., Valentini, G., Fernández-Oto, C., Dorigo, M., & Trianni, V. (2015). A design pattern for decentralised decision making. *PLoS ONE, 10*(10), e0140950. https://doi.org/10.1371/journal.pone.0140950.

Reina, A., Marshall, J. A. R., Trianni, V., & Bose, T. (2017). Model of the best-of-*n* nest-site selection process in honeybees. *Physical Review, 95,* 052411. https://doi.org/10.1103/PhysRevE.95.052411.

Rubenstein, M., Ahler, C., Hoff, N., Cabrera, A., & Nagpal, R. (2014). Kilobot: A low cost robot with scalable operations designed for collective behaviors. *Robotics and Autonomous Systems, 62*(7), 966–975.

Schranz, M., Umlauft, M., Sende, M., & Elmenreich, W. (2020). Swarm robotic behaviors and current applications. *Frontiers in Robotics and AI, 7,* 36. https://doi.org/10.3389/frobt.2020.00036.

Seeley, T. D., & Buhrman, S. C. (2001). Nest-site selection in honey bees: How well do swarms implement the "best-of-n" decision rule? *Behavioral Ecology and Sociobiology, 49,* 416–427.

Stone, M. (1961). The opinion pool. *The Annals of Mathematical Statistics, 32*(4), 1339–1342.

Sumpter, D. J., & Pratt, S. C. (2009). Quorum responses and consensus decision making. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1518), 743–753. https://doi.org/10.1098/rstb.2008.0204.

Valentini, G., Hamann, H., & Dorigo, M. (2015). Efficient decision-making in a self-organizing robot swarm: On the speed versus accuracy trade-off. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015)* (pp. 1305–1314). Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

Valentini, G., Ferrante, E., Hamann, H., & Dorigo, M. (2016). Collective decision with 100 Kilobots: Speed versus accuracy in binary discrimination problems. *Autonomous Agents and Multi-Agent Systems, 30*(3), 553–580. https://doi.org/10.1007/s10458-015-9323-3.

Valentini, G., Ferrante, E., & Dorigo, M. (2017). The best-of-*n* problem in robot swarms: Formalization, state of the art, and novel perspectives. *Frontiers in Robotics and AI, 4, 9*. https://doi.org/10.3389/frobt.2017.00009.